COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 106 (2012) 127–138

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

EPILEPSIAE – A European epilepsy database

Matthias Ihle[a,b,*], Hinnerk Feldwisch-Drentrup[a,b,c,d], César A. Teixeira[e], Adrien Witon[f], Björn Schelter[b,g], Jens Timmer[b,g,h], Andreas Schulze-Bonhage[a,c]

[a] Epilepsy Center, University Hospital Freiburg, Germany
[b] Center for Data Analysis and Modeling (FDM), University of Freiburg, Germany
[c] Bernstein Center for Computational Neuroscience, University of Freiburg, Germany
[d] Department of Neurobiology and Biophysics, Faculty of Biology, University of Freiburg, Germany
[e] Centre for Informatics and Systems (CISUC), University of Coimbra, Portugal
[f] Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (CRICM), INSERM UMRS 975 – CNRS UMR 7225-UPMC, Hôpital de la Pitié-Salpêtrière, Paris, France
[g] Department of Physics, University of Freiburg, Germany
[h] Freiburg Institute for Advanced Studies, University of Freiburg, Germany

ARTICLE INFO

Article history:
Received 1 February 2010
Received in revised form 6 July 2010
Accepted 16 August 2010

Keywords:
Database
Schema
Epilepsy
Seizure prediction
EEG

ABSTRACT

With a worldwide prevalence of about 1%, epilepsy is one of the most common serious brain diseases with profound physical, psychological and, social consequences. Characteristic symptoms are seizures caused by abnormally synchronized neuronal activity that can lead to temporary impairments of motor functions, perception, speech, memory or, consciousness.

The possibility to predict the occurrence of epileptic seizures by monitoring the electroencephalographic activity (EEG) is considered one of the most promising options to establish new therapeutic strategies for the considerable fraction of patients with currently insufficiently controlled seizures.

Here, a database is presented which is part of an EU-funded project "EPILEPSIAE" aiming at the development of seizure prediction algorithms which can monitor the EEG for seizure precursors. High-quality, long-term continuous EEG data, enriched with clinical metadata, which so far have not been available, are managed in this database as a joint effort of epilepsy centers in Portugal (Coimbra), France (Paris) and Germany (Freiburg).

The architecture and the underlying schema are here reported for this database. It was designed for an efficient organization, access and search of the data of 300 epilepsy patients, including high quality long-term EEG recordings, obtained with scalp and intracranial electrodes, as well as derived features and supplementary clinical and imaging data. The organization of this European database will allow for accessibility by a wide spectrum of research groups and may serve as a model for similar databases planned for the future.

* *Corresponding author at*: Epilepsy Center, University Hospital Freiburg, Breisacher Str. 64, 79106 Freiburg, Germany. Tel.: +49 761 270 9313.
E-mail address: matthias.ihle@uniklinik-freiburg.de (M. Ihle).
0169-2607/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2010.08.011

1. Introduction

Epilepsy is one of the most common diseases of the human brain, with a prevalence of more than 3 million patients in Europe alone. Epilepsy is characterized by sudden changes in brain dynamics that lead to abnormal synchronization of extended brain networks, the so-called "seizures". These seizures are characterized by transient impairments of sensation, thinking and motor control. In most patients, seizures are infrequent, occupying much less than 0.1% of the time. Due to their apparently unpredictable occurrence, patients are, however, suffering from restrictions in several domains, e.g., physically due to the risk of trauma, socially due to driving and occupational restrictions, psychologically due to a feeling of helplessness [1,2]. Furthermore, a continuous prophylactic medical treatment is presently being offered that renders patients liable to side effects [3].

One third of the patients presently do not respond to a continuous prophylactic treatment in maximally tolerated dosages. Particularly for this large patient group, new treatment concepts have to be developed. Such concepts could be based on prediction-based warnings to patients or on prediction-based closed-loop interventions. These would change continuous treatment to timely targeted medical or electrical interventions prior to seizures [4].

For this purpose, the EU-funded project "EPILEPSIAE" (www.epilepsiae.eu) aims at the development of EEG-based seizure prediction algorithms that capture changes in electroencephalography (EEG) dynamics and use these for a warning of the patient. Up to now, visual inspection did not reveal precursors for seizures. So far, one of the factors limiting the evaluation of prediction algorithms have been limitations in the quality and duration of long-term EEG data available for a valid evaluation of seizure prediction methods and their performances [5,6], and limited access of groups with knowledge in the field of time series analysis to such EEG data.

Accordingly, both in Europe and in the USA efforts are made to overcome this obstacle for the development of seizure prediction. The EPILEPSIAE project will gather the largest and most comprehensive epilepsy database existing worldwide. It is based on the common effort of three European epilepsy centers (Freiburg, Germany; Paris, France; and Coimbra, Portugal), which contribute EEG data from long-term monitoring of epilepsy patients as well as standardized annotations and clinical metadata. After its completion, it is planned that this database project will offer access to data for research groups throughout the world and will collaborate with a US database that is presently being designed.

We here report on the design and the resulting schema of the database. Additionally, we present client applications and the current status regarding content: the input and management of large data sets.

2. Background

Seizure prediction is based on the identification of precursors in long-term EEG time series. A major concern is access to suitable data for the systematic application and evaluation of seizure prediction methods. Particularly for research groups with no direct contact to epilepsy centers, access to clinical data often imposes severe constraints for the progress of research. This has been pinpointed already in 2001 [7] when the importance of freely available datasets was emphasized: "One recent initiative, the development of open databases that could serve as a repository of clinical data that are difficult and expensive to obtain, provides an emerging strategy that should prove indispensable for testing competing algorithms". This demand for comprehensive databases that can be used for a cross comparison of methods developed on the basis of different, publicly available data sources has led to the publication of some EEG dataset collections in the prediction community, which can be freely accessed.

The Bonn EEG database [8,9] consists of data of five subjects, including different channels per patient, recorded with a sampling rate of 173 Hz. Datasets are discontinuous. For each dataset, about 40 min of EEG are provided.

The Flint Hills Scientific, L.L.C., Public ECoG Database [10] (supported by NIH/NINDS Grant No. 3R01NS046602-03S1) consists of a total of 1419 h of continuous intracranial recordings at 249 Hz for ten patients. Additionally, the database contains meta information about the 59 contained seizures and information about the electrode locations. It provides the EEG from all recording electrodes, which range from 48 to 64 per patient.

The Freiburg EEG database [11] contains invasive long-term EEG recordings of 21 patients, acquired with a sampling rate of 256 Hz obtained during invasive pre-surgical epilepsy monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany. For each patient, the recordings of three focal and three extra-focal electrode contacts are available. In contrast to the other databases, here, a clear separation between ictal and interictal phases for each of the patients is given. For ictal events, files with epileptic seizures and at least 50 min pre-ictal data are provided; the interictal data contain at least 24 h of EEG recordings without seizure activity.

Although this database ranks among the most comprehensive of the currently available EEG databases and is used by more than 180 research groups worldwide, it is clear that there are still drawbacks: first, there is the general lack of long term, continuous recordings. Second, there is only limited information on clinical metadata and annotations.

Overall, all these databases consist of unstructured EEG recordings, supplemented by some clinical information. For example, the correlation between seizures, their origin and propagation and, the localization of electrodes or, other exact information about the brain topography or the epilepsy characteristics are not available when using the data of these databases.

As provider of the Freiburg database, we collect information about the purposes of the database users. The use for research on seizure prediction was most frequently named, followed by seizure detection. But it became apparent that the application domain is much wider. It ranges from detection and analysis of interictal spikes over the automatic classification of EEG signals, for instance through machine learning, to the general application of time series techniques.

Additionally, various direct user requests prove the general demand for metadata about the EEG in addition to the raw recording data, like the type and localization of the